

"Express Mail" mailing label number EV 327174842 US

Date of Deposit: February 12, 2004

Our Case No.10808/123

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
APPLICATION FOR UNITED STATES LETTERS PATENT

INVENTOR:

ACHIM GRATZ
KLAUS KNOBLOCH
FRANZ SCHULER

TITLE:

SEMICONDUCTOR CONTROL
ARRANGEMENT WITH DITCH
ISOLATION AND MANUFACTURE
PROCEDURE

ATTORNEY:

John J. King
Reg. No. 35,918
BRINKS HOFER GILSON & LIONE
P.O. BOX 10395
CHICAGO, ILLINOIS 60610
(312) 321-4200

Beschreibung

Halbleiter-Schaltungsanordnung mit Grabenisolation und Herstellungsverfahren

5

Die Erfindung betrifft eine Halbleiter-Schaltungsanordnung mit einem Substrat, das in der angegebenen Reihenfolge trägt:

- eine dotierte Halbleiterschicht eines ersten Leitfähigkeitstyps bzw. Leitungstyps,
- 10 - eine elektrisch isolierende Schicht,
- und eine elektrisch leitfähige oder eine elektrisch isolierende ladungsspeichernde Schicht, die zur Speicherung von Ladungen geeignet ist.

15 Außerdem enthält die Halbleiter-Schaltungsanordnung mindestens einen Graben, der die ladungsspeichernde Schicht durchdringt und sich auch in die dotierte Halbleiterschicht hinein erstreckt.

20 Das Substrat ist beispielsweise ein Plättchen aus einem Halbleitermaterial, z.B. aus Silizium. Die zur Speicherung von Ladungen geeignete Schicht wird insbesondere bei Schaltungsanordnungen mit Speicherzellen auch als Floating-Gate bezeichnet.

25

Es ist Aufgabe der Erfindung, eine einfach herzustellende und eine einfach anzusteuernde integrierte Schaltungsanordnung anzugeben, die insbesondere sehr gute elektrische Eigenschaften hat. Insbesondere soll eine Schaltungsanordnung mit einer

30 Vielzahl von Speicherzellen angegeben werden. Außerdem betrifft die Erfindung ein Verfahren, mit dem sich insbesondere die Halbleiter-Schaltungsanordnung herstellen lässt.

Die auf die Schaltungsanordnung bezogene Aufgabe wird durch

35 eine Schaltungsanordnung mit den Merkmalen des Patentanspruchs 1 gelöst. Weiterbildungen sind in den Unteransprüchen angegeben.

Die Erfindung geht von der Überlegung aus, dass es grundsätzlich zwei Möglichkeiten für die Herstellung des Grabens gibt. So ist es möglich, den Graben abgesehen von Hilfsschichten, die nach dem Ausbilden des Grabens wieder vollständig entfernt werden, vor dem Aufbringen von Schichten herzustellen, die in der Schaltungsanordnung verbleiben, so dass diese Schichten nicht von dem Graben durchdrungen werden. Andererseits besteht die Möglichkeit, den Graben erst nach dem Aufbringen von in der Schaltungsanordnung verbleibenden Schichten auf das Substrat einzubringen, so dass diese Schichten von dem Graben durchdrungen werden.

Die Erfindung geht weiterhin von der Überlegung aus, dass das Erzeugen des Grabens nach dem Aufbringen von in der Schaltungsanordnung verbleibenden Schichten gleichzeitig die Strukturierung dieser Schichten und die Ausrichtung des Grabens bezüglich der strukturierten Bereiche zulässt, d.h. ein sogenanntes Self-Alignment. Es besteht jedoch die Möglichkeit, dem Graben weitere Funktionen in der integrierten Schaltungsanordnung zuzuordnen. Bei der erfindungsgemäßen Schaltungsanordnung dient der Graben außerdem noch zum Unterteilen der dotierten Halbleiterschicht. Dies erfordert, dass der Graben tiefer ist, als die Dicke der dotierten Halbleiterschicht. Damit hat der Graben bei der erfindungsgemäßen Schaltungsanordnung neben der Isolierfunktion zum Isolieren von benachbarten Bauelementen noch zwei weitere Funktionen, nämlich:

- die Strukturierung der ladungsspeichernden Schicht, und
- die Strukturierung und Isolierung der dotierten Halbleiterschicht.

Bei einer Ausgestaltung enthält die Schaltungsanordnung mehrere nebeneinander angeordnete Gräben, beispielsweise zueinander parallel liegende Gräben. Zwischen den Gräben sind jeweils eine Vielzahl von Speicherzellen angeordnet, insbesondere EEPROM-Speicherzellen oder sogenannte Flash-EEPROM-

Speicherzellen (Electrical Erasable Programmable Read Only Memory). Bei einer nächsten Ausgestaltung ist die ladungsspeichernde Schicht quer zu der Richtung, in der die Gräben liegen, in ladungsspeichernde Bereiche unterteilt.

5

Bei einer Weiterbildung der Schaltungsanordnung ist zwischen der dotierten Halbleiterschicht und dem Substrat eine dotierte Halbleiterschicht mit zum ersten Leitfähigkeitstyp entgegengesetztem Leitfähigkeitstyp angeordnet. Der Graben durchdringt auch die Halbleiterschicht entgegengesetztem Leitungstyps und erstreckt sich bis in das Substrat. Die Unterteilung einer dotierten Halbleiterschicht durch zwei Gräben, z.B. in der sogenannten Bitline-Richtung, führt bereits zu einer Vielzahl von Vorteilen bezüglich des Betriebes der Speicherzellen. Beispielsweise können Speicherzellen selektiv gelöscht werden. Das Abgrenzen einer weiteren unter der oberen Halbleiterschicht liegenden unteren Halbleiterschicht mit Hilfe der Gräben führt zu einer weiteren Verbesserung der elektrischen Eigenschaften. Beispielsweise reduziert sich die Übergangskapazität zwischen den beiden Halbleiterschichten für jede einzelne Bitleitung. Ferner ergeben sich zusätzliche Optionen für die elektrische Ansteuerung der Bauelemente.

Bei einer alternativen Weiterbildung mit zwei zusätzlichen Halbleiterschichten endet der Graben dagegen in der Halbleiterschicht entgegengesetzten Leitungstyps, so dass nur die Halbleiterschicht des ersten Leitungstyps durch die Gräben getrennt wird. Diese Maßnahme ist für viele Anwendungen ausreichend und einfacher durchzuführen als die Durchtrennung beider oder mehr als zwei Halbleiterschichten mit Hilfe der Gräben.

Bei einer Ausgestaltung ist das Substrat ein Halbleitersubstrat, das vorzugsweise Silizium enthält oder aus Silizium besteht. Das Silizium ist beispielsweise schwach vordotiert. Bei einer anderen Ausgestaltung bildet die dotierte Halbleiterschicht das Kanalgebiet eines Transistors bzw. die Kanal-

gebiete einer Vielzahl von Transistoren. Bei einer nächsten Ausgestaltung ist der erste Leitfähigkeitstyp der p-Leitfähigkeitstyp, d.h. einer elektrischen Leitung durch Defektelektronen oder sogenannte Löcher. Bei einer alternativen Ausgestaltung ist der erste Leitfähigkeitstyp der n-Leitfähigkeitstyp, d.h. einer elektrischen Leitung durch Leitungselektronen.

Bei einer nächsten Ausgestaltung enthält die elektrisch isolierende Schicht ein Oxid oder besteht aus einem Oxid, insbesondere aus einem vorzugsweise thermisch erzeugten oder abgeschiedenen Siliziumdioxid. Die an das Substrat grenzende elektrisch isolierende Schicht wird beim Programmieren und Löschen der Speicherzellen von Ladungsträgern durchtunnelt oder überwunden. In der ladungsspeichernden Schicht werden beim Programmieren bspw. Elektronen oder Elektronenlöcher gespeichert. Ladungsträger, die in diesem Zusammenhang auch als hot carrier bezeichnet werden, werden beim Programmieren und/oder Löschen auf Grund eines elektrischen Feldes so beschleunigt, dass sie eine Energiebarriere zwischen der ladungsspeichernden Schicht und dem Substrat überwinden können. Alternativ kann durch Anlegen eines geeigneten Potentialgefälles die Energiebarriere so verringert werden, dass Ladungsträger diese durchtunneln können.

Bei einer nächsten Ausgestaltung enthält die ladungsspeichernde Schicht polykristallines Silizium oder besteht aus polykristallinem Silizium, vorzugsweise aus dotiertem polykristallinen Silizium. Bei einer alternativen Ausgestaltung enthält die ladungsspeichernde Schicht ein Nichtmetallnitrid oder besteht aus einem Nichtmetallnitrid, insbesondere aus Siliziumnitrid. Bei einer anderen alternativen Ausgestaltung enthält die ladungsspeichernde Schicht ein anderes Material, das in der Lage ist, Ladungsträger beispielsweise in Materialfehlstellen zu binden, wie beispielsweise Aluminiumoxid oder Hafniumoxid.

Bei einer Weiterbildung der erfindungsgemäßen Schaltungsanordnung ist eine elektrisch leitfähige Schicht vorgesehen, die als Wortleitungen strukturiert ist. Zwischen der elektrisch leitfähigen Schicht und der ladungsspeichernden Schicht ist eine weitere elektrisch isolierende Schicht angeordnet. Diese elektrisch leitfähige Schicht und diese elektrisch isolierende Schicht werden bei einer Ausgestaltung nicht von dem Graben bzw. von den Gräben durchdrungen, die sich bis in die Halbleiterschicht erstrecken.

Bei einer Ausgestaltung enthält die elektrisch leitfähige Schicht ein polykristallines Material oder ein Metall. Beispielsweise besteht die elektrisch leitfähige Schicht aus polykristallinem Silizium, insbesondere aus dotiertem polykristallinen Silizium. Bei einer nächsten Ausgestaltung ist die elektrisch leitfähige Schicht in Streifen unterteilt, die quer bzw. im Winkel von 90 Grad zu den Gräben liegen. Bei einer anderen Ausgestaltung hat die elektrisch leitfähige Schicht Unterteilungen an Stellen, an denen auch die ladungsspeichernde Schicht unterteilt ist, d.h. dass beide Schichten mit der gleichen Maske strukturiert worden sind.

Bei einer Weiterbildung der Schaltungsanordnung gibt es mindestens einen im Vergleich zu dem die elektrisch isolierende Schicht durchdringenden Graben flacheren und breiteren Graben, der in der Halbleiterschicht des ersten Leitfähigkeitstyps angeordnet ist und der von dem die elektrisch isolierende Schicht durchdringenden tiefen Graben durchdrungen wird. Durch diese Maßnahme entstehen ohne größeren prozesstechnischen Aufwand Freiheitsgrade für die Verfahrensdurchführung, weil der flache Graben als zusätzliche Isolation benutzt werden kann. Der prozesstechnische Mehraufwand ist deshalb gering, weil in vielen Schaltungsanordnungen ohnehin flache Gräben zu erzeugen sind. Insbesondere werden flache Gräben in Logikschaltungen verwendet. Flache Gräben haben typischerweise eine Tiefe kleiner als 500 nm (Nanometer). Dagegen hat der tiefe Graben eine Tiefe größer als 700 nm, größer als 1 μm

(Mikrometer) oder sogar größer als $1,5 \mu\text{m}$. Die Tiefe des tiefen Grabens hängt insbesondere von den Spannungsverhältnissen ab, weil diese wiederum die Dicke der dotierten Halbleiterschichten bestimmen, die durch die tiefen Gräben parallel zu den Bitleitungen unterteilt werden sollen.

Bei einer nächsten Weiterbildung der Schaltungsanordnung durchdringt der flache Graben die ladungsspeichernde Schicht und/oder die elektrisch isolierende Schicht nicht. Damit muss der flache Graben vor dem Aufbringen dieser beiden Schichten hergestellt und insbesondere auch gefüllt worden sein. Dies bedeutet, dass die Prozessierung von flachen Gräben und tiefen Gräben unabhängig voneinander erfolgt. Insbesondere lassen sich die Tiefen der verschiedenen Grabenarten unabhängig voneinander festlegen und optimieren. Weiterhin werden bei dieser Verfahrensführung Probleme vermieden, die aufgrund der starken Höhenunterschiede beim gleichzeitigen Füllen von flachen Gräben und tiefen Gräben beim nachfolgenden Einebnen der Oberfläche entstehen.

Bei einer Weiterbildung ist der flache Graben vollständig mit einem elektrisch isolierenden Material gefüllt oder der flache Graben enthält ein elektrisch isolierendes Material, beispielsweise Siliziumdioxid. Bei einer nächsten Weiterbildung ragt der flache Graben symmetrisch über den tiefen Graben hinaus, so dass die Isolationseigenschaften in mehrere Richtungen gleich gut sind.

Bei einer nächsten Weiterbildung gibt es mindestens einen weiteren flachen Graben, der von keinem Graben durchdrungen ist, insbesondere von keinem tiefen Graben. Bei einer nächsten Weiterbildung liegt der von dem tiefen Graben durchdrungene flache Graben in einem Speicherzellenfeld und der nicht von einem Graben durchdrungene flache Graben in einer Logik-Schaltungsanordnung, in der beispielsweise logische Grundfunktionen erbracht werden, z.B. NAND-Schaltsfunktionen. Die Logikschaltung ist bspw. Teil eines monolithischen Schalt-

kreises, der auch ein Speicherzellenfeld mit tiefen Gräben enthält.

Bei einer nächsten Weiterbildung sind die nicht vom tiefen
5 Graben durchdrungene elektrisch leitfähige Schicht und/oder
die nicht vom tiefen Graben durchdrungene elektrisch isolie-
rende Schicht zumindest teilweise in dem flachen Graben ange-
ordnet. Durch diese Maßnahme kann eine Aussparung in die vom
tiefen Graben nicht durchdrungenen elektrisch leitfähigen
10 Schicht ohne strenge Anforderungen an deren Tiefe eingebracht
werden. Wichtig ist nur, dass die elektrisch leitfähige
Schicht vollständig unterbrochen wird. Für das Hineinragen
der Aussparung in den flachen Graben gibt es einen größeren
Spielraum, von beispielsweise mehr als 10 nm oder mehr als
15 20 nm. Trotz verschiedener Tiefen ist gewährleistet, dass der
Kopplungsfaktor zwischen der Kapazität der ladungsspeichern-
den Schicht und der elektrisch leitfähigen Schicht relativ
unabhängig von der Tiefe ist, wenn die Aussparung innerhalb
der Grabenränder des flachen Grabens liegt.

20 Bei einer nächsten Weiterbildung der Schaltungsanordnung
enthält die Schaltungsanordnung eine weitere ladungsspei-
chernde Schicht, die an die ladungsspeichernde Schicht an-
grenzt und die bei einer Ausgestaltung aus dem gleichen Mate-
25 rial besteht. In der weiteren ladungsspeichernden Schicht ist
mindestens eine Aussparung angeordnet, deren Boden vorzugs-
weise vollständig innerhalb des Randes des tiefen Grabens
bzw. des flachen Grabens liegt. Auch die Tiefe dieser Ausspa-
rung ist unkritisch, sofern die weitere ladungsspeichernde
30 Schicht vollständig strukturiert wird. Hinsichtlich des Kop-
pelfaktors der Kapazitäten gelten die gleichen Zusammenhänge
wie im vorhergehenden Absatz erläutert.

Bei einer anderen Weiterbildung ist der Graben mit einem
35 elektrisch isolierenden Material gefüllt oder der Graben
enthält ein elektrisch isolierendes Material. Zum Füllen des
Grabens sind insbesondere Oxide wie beispielsweise Silizium-

dioxid geeignet. Bei einer nächsten Weiterbildung enthält der Graben ein von der Grabenwand isoliertes elektrisch leitfähiges oder elektrisch halbleitendes Material, beispielsweise ein polykristallines Material, insbesondere polykristallines Silizium, das dotiert oder undotiert ist.

Die Erfindung betrifft außerdem ein Verfahren zum Herstellen einer Halbleiter-Schaltungsanordnung, insbesondere zum Herstellen der erfindungsgemäßen Schaltungsanordnung oder einer ihrer Weiterbildungen. Damit gelten die oben genannten technischen Wirkungen auch für das Verfahren.

Bei einer Weiterbildung des erfindungsgemäßen Verfahrens wird zum Einbringen des tiefen Grabens eine Hartmaskenschicht verwendet. Die Hartmaskenschicht lässt sich vor dem Füllen des Grabens entfernen. Jedoch lässt sich die Hartmaskenschicht auch erst nach dem Füllen des Grabens mit einem Füllmaterial und einem anschließenden Rückätzen des Füllmaterials entfernen. Durch diese Maßnahme wird erreicht, dass beim Rückätzen unter der Hartmaske liegende Schichten durch die Hartmaske geschützt sind.

Im Folgenden werden Ausführungsbeispiele der Erfindung an Hand der beiliegenden Zeichnungen erläutert. Darin zeigen:

Figuren 1A bis 1F

Herstellungsstufen gemäß einer ersten Verfahrensvariante mit einer Hartmaskenschicht, die unmittelbar nach dem Herstellen von tiefen Gräben entfernt wird,

Figur 2 eine Herstellungsstufe gemäß einer zweiten Verfahrensvariante mit einer Hartmaskenschicht, die über mehrere Verfahrensschritte genutzt wird,

Figuren 3A und 3B

Herstellungsstufen gemäß einer dritten Verfahrensvariante mit flachen Gräben, die von tiefen Gräben durchdrungen werden, und

Figuren 4A und 4B

Herstellungsstufen gemäß einer weiteren Verfahrensvariante mit Floating-Gate-Doppelschicht.

5 Figur 1A zeigt ein schwach p-dotiertes Halbleitersubstrat 10 aus Silizium. In das Halbleitersubstrat 10 wurde eine n-dotierte Halbleiterschicht eindotiert, die beispielsweise in einer Tiefe von 800 nm bis 1,6 μm (Mikrometer) liegt. Außerdem wurde im Halbleitersubstrat 10 eine p-dotierte Halbleiterschicht 14 erzeugt, die sich von der Oberfläche des Halbleitersubstrats bis in eine Tiefe von etwa 800 nm erstreckt. An den Rändern eines Zellenfeldes können die Halbleiterschichten 12 und 14 als Wanne ausgebildet sein, d.h. dass sie bis an die Oberfläche des Halbleitersubstrats 10 geführt sind. Jedoch werden bei einem anderen Ausführungsbeispiel auch andere Kontaktierungsmöglichkeiten genutzt.

Beispielsweise werden die Dotierungen der Halbleiterschichten 12 und 14 durch Implantation hergestellt. Wannenförmig ausgebildete Halbleiterschichten 12 und 14 werden auch als n-well oder als p-well bezeichnet.

Anschließend wird beispielsweise thermisch eine Oxidschicht 16 aufgebracht, die beispielsweise eine Dicke von 6 bis 15 nm hat. Auf der Oxidschicht 16 wird dann eine Floating-Gate-Schicht 18 abgeschieden, die beispielsweise aus in-situ-dotiertem polykristallinen Silizium besteht und beispielsweise eine Dicke von 50 bis 150 nm hat.

30 In einem folgenden Verfahrensschritt wird auf der Floating-Gate-Schicht 18 eine Hartmaskenschicht 20 aufgebracht, beispielsweise eine TEOS-Schicht (Tetra Ethyl Ortho Silicate) mit einer Dicke von mehreren 100 nm. Die Dicke hängt von der Selektivität der Grabenätzung ab. D.h. bei einer sehr selektiven Ätzung oder einem anderen Hartmaskenmaterial kann die Dicke u.U. auch dünner sein.

Wie in Figur 1B dargestellt, wird zunächst die Hartmaskenschicht 20 mit Hilfe eines Lithografieverfahrens und einer nicht dargestellten Fotomaske strukturiert, wobei zunächst nur in der Hartmaskenschicht 20 eine Aussparung 30 entsteht. Beim Herstellen der Aussparung 30 wird beispielsweise zeitgesteuert geätzt. Nach dem Strukturieren der Hartmaskenschicht 20 wird der Fotolack entfernt. Anschließend wird mit Hilfe der strukturierten Hartmaskenschicht 20 die Aussparung 30 zu einem Graben 32 erweitert, der eine Tiefe von beispielsweise etwa 1,8 μm hat, gemessen von der Grenze zwischen Hartmaskenschicht 20 und Floating-Gate-Schicht 18. Zum Ätzen des Grabens 32 wird beispielsweise ein reaktives Ionenätzen RIE (Reactive Ion Etching) durchgeführt. Der Graben 32 hat beispielsweise eine Breite von 200 nm. Beim Ätzen des Grabens 32 wird die Hartmaskenschicht 20 bereits stark gedünnt, so dass nur noch eine Ätzreserve von beispielsweise 100 nm verbleibt.

Wie in Figur 1C dargestellt, wird anschließend die Hartmaskenschicht 20 entfernt bzw. weggeätzt. Anschließend wird eine sogenannte Liner-Oxidation durchgeführt, bei der an der Wand des Grabens 32 und auf der Floating-Gate-Schicht 18 eine Oxidschicht 40 erzeugt wird, die beispielsweise eine Dicke von 40 nm hat. Nach dem Erzeugen der Liner-Oxidschicht 40 wird eine polykristalline Siliziumschicht 42 abgeschieden, die den Graben 32 vollständig füllt. Zum Abscheiden der Siliziumschicht 42 wird beispielsweise ein Niederdruck-CVD-Verfahren (Chemical Vapor Deposition) genutzt. Im Ausführungsbeispiel ist die polykristalline Siliziumschicht 42 undotiert. Bei einem anderen Ausführungsbeispiel wird jedoch eine dotierte polykristalline Siliziumschicht 42 hergestellt.

Wie in Figur 1D dargestellt, werden anschließend außerhalb des Grabens 32 liegende Bereiche der Siliziumschicht 42 entfernt, beispielsweise mit Hilfe eines reaktiven Ionenätzverfahrens. Außerdem wird im oberen Bereich des Grabens 32 selektiv zur Oxidschicht 40 die Siliziumschicht 42 entfernt, beispielsweise bis in eine Tiefe von 300 nm unterhalb der

Grenze zwischen Oxidschicht 16 und p-dotierter Halbleiterschicht 14. Nach dem Ätzen der Siliziumschicht 42 wird ein isolierende Füllmaterial 50 in den oberen Teil des Grabens 32 eingebracht, beispielsweise Siliziumdioxid mit Hilfe eines HDP-Verfahrens (High Density Plasma). Im oberen Bereich des Grabens 32 und auf der Floating-Gate-Schicht 18 entsteht eine Oxidschicht 50.

Wie in Figur 1E dargestellt, wird anschließend die Oxidschicht 50 ganzflächig zurückgeätzt, wobei eine Aussparung 60 im oberen Bereich des Grabens 32 entsteht. Das Rückätzen der Oxidschicht 50 wird beispielsweise mit Hilfe eines RIE-Verfahrens (Reactive Ion Etching) oder mit Hilfe einer Nassätzung durchgeführt. Der Boden der Aussparung 60 sollte nicht unterhalb der Grenze zwischen Oxidschicht 16 und p-dotierter Halbleiterschicht 14 liegen.

Wie in Figur 1F gezeigt, wird anschließend eine Dielektrikumsschicht 70 abgeschieden, beispielsweise eine ONO-Schicht (Oxid-Nitrid-Oxid). Die Dielektrikumsschicht 70 hat beispielsweise eine Dicke von weniger als 20 nm. Nach dem Aufbringen der Dielektrikumsschicht 70 wird eine Control-Gate-Schicht 72 aufgebracht, beispielsweise aus in-situ-dotiertem polykristallinen Silizium und mit einer Dicke von bspw. größer 100 nm.

In nicht dargestellten folgenden Verfahrensschritten werden die Control-Gate-Schicht 72, die Dielektrikumsschicht 70 und die Floating-Gate-Schicht 18 gleichzeitig in einer Wordline-Richtung strukturiert, die parallel zur Blattebene liegt, siehe Pfeil 74. Eine Bitline-Richtung liegt senkrecht zur Blattebene und stimmt mit der Richtung des Grabens 32 überein. Nach dem Strukturieren der Control-Gate-Schicht 72 werden in die p-dotierte Halbleiterschicht 14 Kanal- und Sourcegebiete eindotiert, die bezogen auf Figur 1F vor bzw. hinter der Blattebene liegen. In weiteren Verfahrensschritten wird eine oder mehrere Metallisierungsschichten zum Kontaktieren

der Speicherzellen aufgebracht. Schließlich wird ein Speicherschaltkreis 76 fertiggestellt, der die in Figur 1F dargestellte Anordnung enthält.

5 Figur 2 zeigt eine Herstellungsstufe gemäß einem zweiten Verfahren, bei dem eine Hartmaskenschicht 20a über mehrere Verfahrensschritte verwendet wird. Ausgehend von einem Halbleitersubstrat 10a werden die oben an Hand der Figuren 1A und 1B aufgeführten Verfahrensschritte zum Erzeugen einer n-dotierten Halbleiterschicht 12a, einer p-dotierten Halbleiterschicht 14a, einer Oxidschicht 16a und einer Floating-Gate-Schicht 18a ausgeführt. Bezüglich der Einzelheiten wird auf die Figuren 1A bis 1B verwiesen. Anschließend wird die Hartmaskenschicht 20a aufgebracht und mit Hilfe eines fotolithografischen Verfahrens strukturiert, wobei die Floating-Gate-Schicht 18a zunächst unstrukturiert bleibt. Nach dem Entfernen des Fotolacks wird die strukturierte Hartmaskenschicht 20a zum Erzeugen eines Grabens 32a verwendet, der die gleichen Eigenschaften wie der Graben 32 hat.

20 Bei noch auf der Oxidschicht 18a vorhandener Hartmaskenschicht 20a wird anschließend eine beispielsweise thermische Liner-Oxidation zum Erzeugen einer Oxidschicht 40a durchgeführt, die an den Wänden des Grabens 32a und auf der Hartmaskenschicht 20a liegt.

In einem nächsten Verfahrensschritt wird eine dotierte oder undotierte polykristalline Siliziumschicht 42a abgeschieden, die den Graben 32a vollständig füllt. Danach wird die polykristalline Siliziumschicht 42a beispielsweise mit Hilfe eines reaktiven Ionenätzverfahrens zurückgeätzt, wobei die Siliziumschicht 42a außerhalb des Grabens 32a und in dessen oberem Bereich entfernt wird. Auch die auf der Hartmaskenschicht 20a liegenden Bereiche der Oxidschicht 20a werden dabei entfernt. Die Hartmaskenschicht 20a schützt die Floating-Gate-Schicht 18a während des Rückätzens.

Nach dem Rückätzen wird die Hartmaskenschicht 20a entfernt. Anschließend wird, so wie oben an Hand der Figuren 1D bis 1F erläutert, weiter prozessiert, d.h. Aufbringen einer der Oxidschicht 50 entsprechenden Oxidschicht usw.

5

Die Figuren 3A und 3B zeigen Herstellungsstufen gemäß einer dritten Verfahrensvariante, bei der ein flacher Graben 100 von einem tiefen Graben 32b durchdrungen wird. Der flache Graben 100 wird in ein dem Halbleitersubstrat 10 entsprechenden Halbleitersubstrat 10b vor oder nach dem Aufbringen einer der Oxidschicht 16 entsprechenden Oxidschicht 16b und einer der Floating-Gate-Schicht 18 entsprechenden Floating-Gate-Schicht 18b erzeugt, beispielsweise vor der Implantation zum Erzeugen einer n-dotierten Halbleiterschicht 12b bzw. einer p-dotierten Halbleiterschicht 14b. Bis auf das Einbringen des Grabens 100 werden die an Hand der Figuren 1A bis 1D erläuterten Verfahrensschritte unverändert ausgeführt, siehe gestrichelte Linie 100 in diesen Figuren. Wenn der in Figur 1D dargestellte Stand erreicht ist, wird eine der Oxidschicht 50 entsprechende Oxidschicht zurückgeätzt, wobei eine Aussparung 60b im oberen Bereich des Grabens 32b entsteht. Beim Zurückätzen ist es unkritisch, wenn der Boden der Aussparung 60b unterhalb der Grenze zwischen der Oxidschicht 16b und der Halbleiterschicht 14b liegt. Beispielsweise kann mehrere 10 nm in den flachen Graben 100 hineingeätzt werden, siehe gestrichelte Linie 102. Der Graben 100 ist mit einem Isoliermaterial gefüllt, beispielsweise mit Siliziumdioxid. Dieses Isoliermaterial bietet auch bei tieferliegendem Boden der Aussparung 60b eine ausreichende Isolierung zwischen dem später aufgebrauchten Control-Gate und der Halbleiterschicht 14b.

Der Graben 100 hat nämlich eine größere Breite als der Graben 32b. Im Ausführungsbeispiel hat der flache Graben 100 eine Breite von 300 nm und eine Tiefe von 400 nm. Bei symmetrischer Anordnung des tiefen Grabens 32b bezüglich des flachen Grabens 100 ragt der flache Graben 100 auf jeder Seite mit

5 einem Abstand A von 50 nm über den tiefen Graben 32b hinaus.
auf Grund der größeren Breite des Grabens 100 werden auch
Anätzungen des Tunneloxides 16b in Randbereichen des Grabens
100 beim Ätzen der Aussparung 60b vermieden, siehe Bereiche
104 und 106, was zu einer höheren Zuverlässigkeit führt.

Wie in Figur 3B dargestellt, wird anschließend eine Die-
lektrikumsschicht 70b aufgebracht, die der Dielektrikum-
schicht 70 entspricht. Danach wird eine Control-Gate-Schicht
10 72b aufgebracht, die der Control-Gate-Schicht 72 entspricht.
Anschließend werden die anderen an Hand der Figur 1F erläu-
terten Verfahrensschritte ausgeführt.

Das Verfahren mit einem flachen Graben, der von einem tiefen
15 Graben durchdrungen wird, wird gemäß einer vierten Verfah-
rensvariante auch bei der an Hand der Figur 2 erläuterten
Verfahrensvariante durchführen, siehe gestrichelte Linie 100
in Figur 2. Das heißt, dass eine der Hartmaskenschicht 20a
entsprechende Hartmaskenschicht auch dann für mehrere Verfah-
20 rensschritte genutzt werden kann, wenn der flache Graben 100
von einem dem tiefen Graben 32a entsprechenden tiefen Graben
durchdrungen wird.

Die Figuren 4A bis 4B zeigen eine fünfte Verfahrensvariante,
25 bei der eine Floating-Gate-Doppelschicht aus einer der Floa-
ting-Gate-Schicht 18 entsprechenden Floating-Gate-Schicht 18c
und einer Floating-Gate-Schicht 110 hergestellt wird. Die
Verfahrensschritte bis zum Rückätzen einer der Oxidschicht 50
entsprechenden Oxidschicht 50c sind die gleichen, wie oben an
30 Hand der Figuren 1A bis 1D erläutert. Für ein Halbleitersub-
strat 10c, eine n-dotierte Halbleiter-Schicht 12c, eine p-
dotierte Halbleiterschicht 14c, eine Oxidschicht 16c, einen
Graben 32c, eine Oxidschicht 40c und eine polykristalline
Siliziumgrabenfüllung 42c wird auf die Ausführungen zu den
35 Figuren 1A bis 1D verwiesen.

Abweichend von dem an Hand der Figur 1D erläuterten Verfahren wird die Oxidschicht 50c nur bis zur Floating-Gate-Schicht 18c zurückgeätzt, wobei die Floating-Gate-Schicht 18c als Ätzstoppschicht dient. Der Graben 32c bleibt also auch in
5 seinem oberen Bereich mit der Oxidschicht 50c gefüllt.

Durch diese Vorgehensweise wird, wie auch an Hand der Figur 3A erläutert, das dünne Oxid der Oxidschicht 16c an den Rändern 112 und 114 des Grabens 32c beim Rückätzen der Oxid-
10 schicht nicht geschädigt.

Danach wird die Floating-Gate-Schicht 110 abgeschieden, beispielsweise polykristallines Silizium, das in-situ dotiert wird.

15

Wie in Figur 4B dargestellt, wird danach mit Hilfe eines fotolithografischen Verfahrens die Floating-Gate-Schicht 110 strukturiert. Dabei wird oberhalb des Grabens 32c eine Aussparung 120 erzeugt, deren Breite geringer als die Breite des
20 Grabens 32c ist. Die Aussparung 120 wird symmetrisch zur Grabenmitte des Grabens 32c ausgerichtet. Beim Ätzen der Aussparung 120 muss nur darauf geachtet werden, dass die Floating-Gate-Schicht 110 vollständig durchtrennt wird. Eine Überätzung ist unkritisch, weil der Boden der Aussparung 120
25 allseitig von dem im Graben 32c verbliebenen Teil der Oxidschicht 50c umgeben ist, siehe gestrichelte Linie 122. Auch bei unterschiedlicher Lage des Bodens der Aussparung 120 verändert sich ein Koppelfaktor der Kapazitäten zwischen dem Floating-Gate und dem Control-Gate der herzustellenden Spei-
30 cherzellen auf Grund des Überstandes der Floating-Gate-Schicht 110 über die Floating-Gate-Schicht 18c nur unwesentlich.

Anschließend werden die oben an Hand der Figur 1F erläuterten
35 Verfahrensschritte durchgeführt. An Stelle einer dreilagigen ONO-Schicht lässt sich auch eine einlagige dielektrische Schicht verwenden.

Gemäß einer sechsten Verfahrensvariante werden die Verfahren gemäß Figur 2 und gemäß der Figuren 4A und 4B kombiniert, so dass die Hartmaske für mehrere Verfahrensschritte auch bei
5 einem Verfahren genutzt wird, bei dem eine Floating-Gate-Doppelschicht erzeugt wird. Das Rückätzen der Oxidschicht wird dann bspw. zeitgesteuert durchgeführt. Danach wird die Hartmaske entfernt.

10 Die Aussparung 120 lässt sich auch breiter als der Graben 32c ausführen. Auf Grund der beiden Lithografieverfahren für die beiden Floating-Gate-Schichten 18c und 110 lassen sich die Breite des Grabens 32c und die Breite der Aussparung 120 unabhängig voneinander wählen.

15 Bei anderen Ausführungsbeispielen wird an Stelle eines Rückätzens auch ein CMP-Verfahren (Chemical Mechanical Polishing) eingesetzt. An Stelle einer Hartmaskenschicht aus TEOS lässt sich auch eine Hartmaskenschicht aus einem anderen Material
20 einsetzen, beispielsweise aus einem Nitrid wie Siliziumnitrid. Falls der tiefe Graben eine geringere Tiefe als in den Ausführungsbeispielen hat, kann an Stelle der Hartmaske auch eine Fotolackschicht verwendet werden.

25 Somit sind mehrere Varianten zum Herstellen von nicht-flüchtigen Speicherzellen erläutert worden, bei denen:

- tiefe Isoliergräben für Flash-EEPROM-Speicherzellen bzw. für einfache EEPROM-Speicherzellen eingesetzt werden,
- sich Verfahrensabschnitte, d.h. sogenannte Module, zum
30 Herstellen von flachen Isolationsgräben (STI - Shallow Trench Isolation) und Module zum Herstellen und Füllen von tiefen Gräben (DTI - Deep Trench Isolation) in einem Herstellungsprozess auf einfache Art insbesondere nacheinander durchführen lassen, und
- 35 - die Probleme der "klassischen" Integration, bei der tiefe Gräben und flache Gräben am Anfang des Verfahrens hergestellt werden, vermieden werden. Insbesondere entstehen

keine Probleme beim Planarisieren von HDP-Oxidfüllungen (High Density Plasma), die eine unterschiedliche Höhe oberhalb der flachen Gräben und der tiefen Gräben haben.

- 5 Bei den erläuterten Ausführungsbeispielen führen tiefe Grabenisolationen zu isolierten Wannenstreifen. Die Gräben im Speicherzellenfeld haben eine größere Tiefe als Gräben in der Ansteuerschaltung des Speicherfeldes oder in einer auf demselben Chip aufgebrachten Logikschaltung. Aufgrund der Füllung mit polykristallinem Silizium lassen sich tiefe Gräben mit einem kleinen Verhältnis von Breite zu Tiefe herstellen, z.B. mit Verhältnissen kleiner 1/4 oder 1/10.

15 Die erläuterten Verfahren sind insbesondere für Speicherzellen geeignet, die auf einem Zellenkonzept beruhen, bei denen die Ladungen aufgrund des Fowler-Nordheim-Tunneleffektes auf das Floating-Gate gebracht werden bzw. vom Floating-Gate entfernt werden.

- 20 Die Variante mit einem tiefen Graben, der keinen flachen Graben durchdringt, ermöglicht es:
- mit einer geringen Chipfläche für die Isolation auszukommen,
 - Probleme der Ausrichtung zwischen flachen Gräben und tiefen Gräben zu vermeiden, so dass keine zusätzlichen Toleranzen vorgesehen werden müssen, und
 - außerhalb des Zellenfeldes liegende flache Gräben problemlos zu planarisieren.

- 30 Die Variante mit einem tiefen Graben, der einen flachen Graben durchdringt, ermöglicht es
- die Isolation zwischen Control-Gate und Substrat bzw. p-dotiertem Bereich zu verbessern,
 - größere Kopplungsfaktoren auf Grund der geringeren Kopplung von Floating-Gate und Substrat zu erzielen,
 - den Kopplungsfaktor besser einstellen zu können, und

- das Fülloxid des Grabens auf leichtere Art rückätzen zu können.

Beide Varianten ermöglichen es:

- 5 - zusätzliche CMP-Schritte (Chemical Mechanical Polishing) zu vermeiden,
- ein CMP-Verfahren mit starken Höhendifferenzen zu vermeiden, und
- 10 - Prozesse zum Herstellen der flachen Gräben von Prozessen zum Herstellen der tiefen Gräben unabhängig voneinander durchzuführen und zu optimieren.

Durch die erläuterten Verfahren lassen sich insbesondere Schädigungen des empfindlichen Tunneloxides insbesondere an
15 den Rändern der Gräben vermeiden bzw. reduzieren, so dass die Ausbeute und Zuverlässigkeit steigt. Außerdem lassen sich so auf einfache Art und Weise in Bitline-Richtung verlaufende Streifen aus den dotierten Wannen erzeugen.

- 20 Bei anderen Ausführungsbeispielen werden die gleichen Strukturen basierend auf anderen Dotierschichtenfolgen hergestellt, z.B. p-Substrat und n-Wanne. Alternativ wird mit n-Substrat und p-Wanne gearbeitet. Bei einer dritten Alternative wird mit n-Substrat und p-Wanne sowie n-Wanne gearbeitet.